Field Crops Research 234 (2019) 55–65

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S

Yan Li[a,b,*], Kaiyu Guan[b,c,*], Albert Yu[d], Bin Peng[b,c], Lei Zhao[c,e,f], Bo Li[d], Jian Peng[g]

[a] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, China
[b] Department of Natural Resources and Environmental Sciences, College of Agriculture, Consumer, and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[c] National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[d] Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[e] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[f] Program in Science, Technology and Environmental Policy (STEP), Princeton University, Princeton, NJ, USA
[g] Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA

ARTICLE INFO

Keywords:
Statistical model
Yield prediction
Yield forecast
Corn
Agriculture

ABSTRACT

Statistical crop models have been a major tool in identifying critical drivers of crop yield, forecasting short-term crop yield, and assessing long-term climate change impacts on agricultural productivity. However, few studies focus specifically on fundamental issues encountered in developing a high-performance statistical crop model for yield prediction. Such issues include: how to select predictors and fitting functions, how to effectively address the spatiotemporal scale issue, weather it is beneficial to include satellite data as explanatory variables, and how to reconcile different model evaluation procedures. In this study, we present our statistical modeling practices for predicting rainfed corn yield in the Midwest U.S. and address the aforementioned issues through comprehensive diagnostic analysis. Our results show that vapor pressure deficit and precipitation at a monthly scale, in spline form with customized knots, define the "Best Climate-only" model among alternative climate variables (e.g., air temperature) and fitting functions (e.g., linear or polynomial), with an out-of-sample (leave-one-year-out) median R^2 of 0.79 and RMSE of 1.04 t/ha (16.6 bu/acre) from 2003 to 2016. Satellite variables, such as MODIS land surface temperature and Enhanced Vegetation Index (EVI), when used as predictors alone, reduce the model's RMSE to 0.93 t/ha (14.8 bu/acre). Adding satellite variables (i.e., EVI in polynomial form) to the "Best Climate-only" model gives the "Best Climate + EVI" model, which has the highest prediction performance of this study, with a median R^2 of 0.85 and RMSE of 0.90 t/ha (14.3 bu/acre). Such a model trained using all data (so-called "global model") in most cases leads to better predictions than the state-specific trained models. However, the global model's prediction performance exhibits considerable regional and interannual variations. The regional-varying performance is related to states' spatiotemporal variability in yield, where states with larger spatial yield variability show higher R^2, and states with smaller temporal yield variability show lower RMSE. Interannual variations in prediction performance are linked to yield variability and degree of wetness, with higher R^2 in years with larger yield variability but increasingly larger RMSE toward wetter years and extreme dry years. These identified spatial and temporal variations of model's performance, together with inconsistent evaluation practices undermine the comparability between statistical modeling studies. Alleviating such comparability issues requires more transparency and open data practices. The statistical model presented in this study provides a benchmark for further development and can be applied to future research related to yield prediction or assessment of climate change impact.

1. Introduction

Accurately predicting crop yield for crop production monitoring and climate change impact assessment is a challenging issue and a key research area in agricultural studies. Two groups of modeling approaches have been widely used to address this problem. The first group is

* Corresponding authors at: Department of Natural Resources and Environmental Sciences, College of Agriculture, Consumer, and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA.
E-mail addresses: yanli.geo@gmail.com (Y. Li), kaiyug@illinois.edu (K. Guan).

https://doi.org/10.1016/j.fcr.2019.02.005
Received 15 June 2018; Received in revised form 31 January 2019; Accepted 5 February 2019
Available online 19 February 2019
0378-4290/ © 2019 Elsevier B.V. All rights reserved.

process-based crop models, which dynamically simulate crop growth and yield formation processes (Bassu et al., 2014; Guan et al., 2017a; Jones et al., 2017; Lobell and Burke, 2010; Müller et al., 2017; Peng et al., 2018a; Rosenzweig et al., 2014). The other group is statistical models, which relate yield to various predictors (e.g., climate) based on empirical relationships derived from measured or observed historical data (Gornott and Wechsung, 2016; Kern et al., 2018; Schlenker and Lobell, 2010). Both groups of models have their pros and cons. Compared with process-based models, the strengths of statistical models lie in their simplicity, fewer inputs required, and relatively high predictive power when sufficient training data are available. Statistical models also have limitations, such as collinearity between predictors, out of sample extrapolation issues (Jones et al., 2017), and data-dependency of prediction performance (Shi et al., 2013). However, statistical models remain as a major tool for yield prediction, and how to build a statistical model with high prediction performance while minimizing the aforementioned concerns is a fundamental question.

Statistical models have been widely applied to understand the crop yield response to climate (Holzkämper et al., 2012; Tack et al., 2015) and to make yield predictions and forecasts[1] (Bussay et al., 2015; Mathieu and Aires, 2016; Peng et al., 2018b; Schauberger et al., 2017). Different forms of statistical models with varying predictors and formulas have been developed to specific scientific questions, such as predicting future yield changes (Lobell et al., 2008; Urban et al., 2012), quantifying effects of warming (Tack et al., 2015) and extreme wetness (Urban et al., 2015), identifying nonlinear temperature responses (Schlenker and Roberts, 2009), and assessing effectiveness of crop adaptation options (Butler and Huybers, 2012). However, most of these models are not designed and optimized to guarantee a high prediction performance. In fact, a systematic investigation of how to build a statistical model with high predictive performance is still lacking (Lobell and Burke, 2010). The associated fundamental questions, such as predictor and fitting function selection, the choice of spatial and temporal scale to train the model, the benefit of using satellite data as explanatory variables, and the model evaluation, are all recurring issues that are worth revisiting. Moreover, it is poorly understood what factors drive the spatial and temporal variations in the model's predictive performance.

The rapid development of remote sensing techniques greatly benefits statistical models in many aspects, particularly for short-term yield forecast (Peng et al., 2018b). Satellite data provide critical information of crop growth and yield (Guan et al., 2017b), such as aboveground biomass or leaf area index (Jin et al., 2017), photosynthetic activities (Guan et al., 2016; Guanter et al., 2014), water stress (Anderson et al., 2011), and phenology (Gao et al., 2017; Zeng et al., 2016). They also enable yield estimation/prediction to be made at very high spatial resolution (Jin et al., 2017; Johnson, 2014; Sakamoto et al., 2014), which was not possible for traditional models relying on climate data (e.g., Schlenker and Roberts, 2009, 2006). Vegetation indices, such as Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), provide vital information on crop biomass and productivity (Guan et al., 2018; Johnson, 2014), and thus have been widely used in crop monitoring and yield estimation (Johnson, 2014; Kern et al., 2018; Sakamoto et al., 2014). Satellite-based land surface temperature (LST) may better capture heat stress (Siebert et al., 2014) and drought impact on crop growth (Karnieli et al., 2010). It is argued that LST can be an alternative to air temperature in data-limited regions, leading to equivalent or even better yield predictions, for example, in Africa (Heft-Neal et al., 2017). This shows that remote sensing data offer a new opportunity to further improve the prediction capability of statistical crop yield models. However, the amount of added value, and how they improve yield prediction of statistical models are still unclear and need to be demonstrated.

In this study, we present our practices for developing a statistical model aimed at high predictive performance, and we also provide a comprehensive diagnostic analysis to advance our understanding of the predictability of statistical models. We used rainfed corn in the U.S. to demonstrate the model's development, but the methodology is adaptable to other crop types. Specifically, we focus on the following questions regarding a high predictive performance model:

1) Which of the three climate variables (i.e., vapor pressure deficit, air temperature, precipitation) and what time scale (monthly or growing season aggregated) are the best predictors for rainfed corn yield?
2) Which fitting function (i.e., linear, polynomial, spline) should be used to effectively capture the climate-yield relationship?
3) How much benefit can satellite data (e.g., EVI and LST) bring to improve the statistical model's performance?
4) At which spatial level should we train the model (a global model or state-specific model)?
5) Why does the model's predictive performance vary across states and years?

2. Data and methods

2.1. Data

The modeling work focuses on the rainfed corn from 12 major states of the U.S. Corn Belt (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin). These 12 states together accounted for 89% of U.S. corn production in 2016. The county-level corn grain yield and harvested area data from 1981 to 2016 were obtained from the National Agricultural Statistics Services (NASS, https://quickstats.nass.usda.gov/). For counties without any irrigated yield records, their yield is considered to be rainfed. For irrigated counties, we only included yield records that are explicitly reported as "nonirrigated" as rainfed yield. This resulted in a total of 32,058 county-year records from 1,051 counties. The original unit of yield bushels/acre was converted to t/ha by multiplying a conversion factor of 0.0628. Yield anomaly (i.e., detrended yield) was calculated by subtracting a global linear yield trend (estimated using data from all counties in all years) from the county yield.

Climate data used in this study were from the Parameter-elevation Relationships on Independent Slopes Model (PRISM, ftp://prism.oregonstate.edu/)(Daly et al., 2008), which include monthly mean vapor pressure deficit (VPD), mean air temperature (T), and monthly precipitation (P) from 1981 to 2016 for the growing season months (May to August). These climate variables were averaged in each county from their 4 km gridded resolution if the center of the 4 km grid is within the county boundary. The mean VPD and air temperature were calculated as an average of maximum and minimum values. Although this is not the best statistical approximation for mean temperature and could have biases for climate change detection studies (Wang, 2014), we don't expect this practice to be an issue for our statistical modeling purpose.

Satellite data were from Moderate Resolution Imaging Spectroradiometer (MODIS, Collection 6), including the 16-day Enhanced vegetation index (EVI, MOD13Q1) from 2001 to 2016 and 8-day daytime land surface temperature (LST, MYD11A2) from 2003 to 2016, at spatial resolutions of 250 m and 1 km, respectively. The daytime MODIS LST from Aqua has an overpass time of 1:30 PM (local time) which approximates the maximum temperature of a day. For each 16-day (8-day) image, we extracted EVI (LST) values from corn pixels identified by crop data layer (CDL) data and then averaged them to county mean values. This process was done on Google Earth Engine. The resulting 16-day (8-day) time series of county mean EVI (LST) was

[1] Yield prediction is a more general term while yield forecast is a more specific term referring to short-term or in-season yield prediction.

Table 1
Summary of climate and satellite predictors used in the model.

Group	Feature	Predictor	Variable specification	Fitting function
Climate variables	Thermal	VPD	$VPD_{May} + VPD_{June} + VPD_{July} + VPD_{Aug}$	Linear, Polynomial (2-order), Spline (1-order)
		T	$T_{May} + T_{June} + T_{July} + T_{Aug}$	
		T_{gs}	$T_{growing\ season\ mean\ from\ May\ to\ Aug}$	
	Water	P	$P_{May} + P_{June} + P_{July} + P_{Aug}$	
		P_{gs}	$P_{growing\ season\ mean\ from\ May\ to\ Aug}$	
Satellite variables	Thermal	LST	$LST_{May} + LST_{June} + LST_{July} + LST_{Aug}$	
	Crop biomass	EVI	$EVI_{May} + EVI_{June} + EVI_{July} + EVI_{Aug}$	

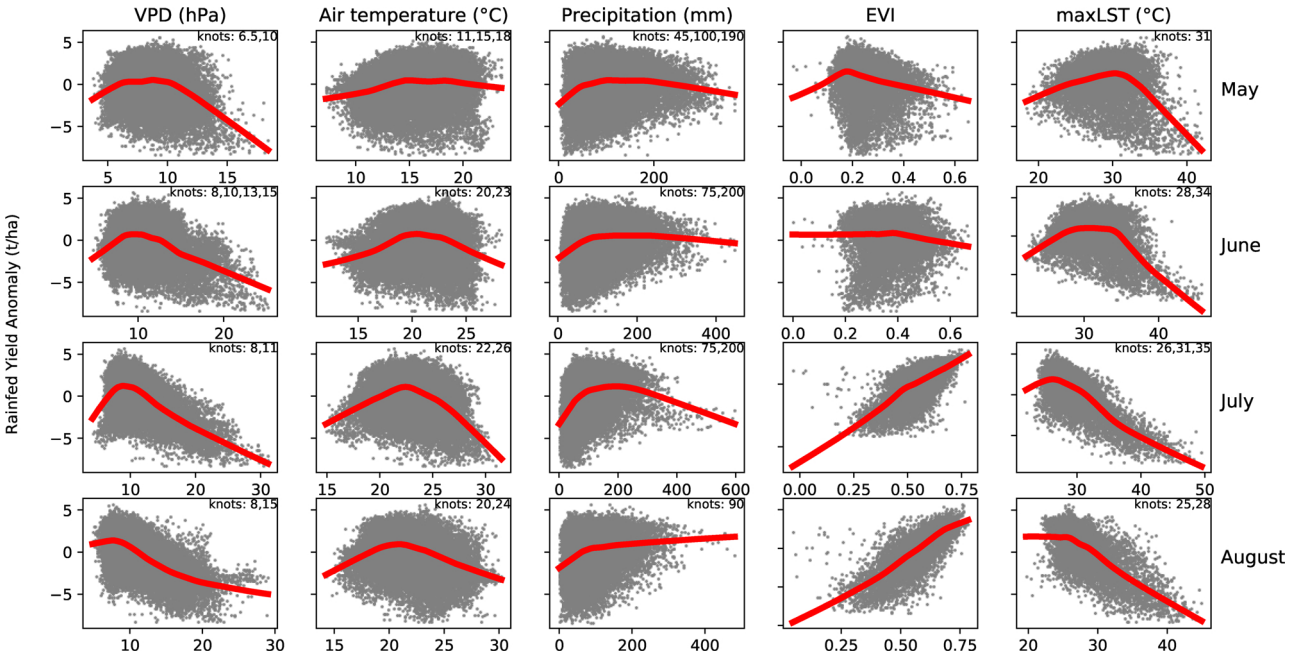

Fig. 1. Scatter plot between yield anomaly and climate/satellite predictors (column-wise) in different months (row-wise). Climate predictors include mean vapor pressure deficit (VPD), mean air temperature, and precipitation. Satellite predictors include EVI and maximum LST from MODIS. The solid red line is the LOWESS fitted line (Seaborn package in Python). The knot values on the upper right corner were determined from the LOWESS fitted line and were used in the spline function. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

further averaged to monthly means for each county. The county monthly mean EVI/LST values for growing season months (May to August) were used as yield predictors. Because the year when CDL data became available differs in different states, some states had no EVI/LST data in early years (before 2007).

2.2. Model configuration

The model in this study predicts yield anomaly (i.e., raw yield after detrending) at the county-level using climate variables (VPD, temperature, precipitation), satellite variables (EVI, LST), and county-specific fixed effects. These climate and satellite variables feature different aspects of crop growth, including thermal and water conditions, and biomass (Table 1). To avoid information redundancy, for each feature, only one predictor was selected to enter the model, except for interaction terms. For instance, among VPD, air temperature, and LST, only one of them can enter the model because they all represent thermal condition and carry similar information (they are highly correlated). Since precipitation and EVI represent water condition and crop biomass information differently, they can enter the model simultaneously. The model structure can be expressed as:

$$Yield\ \ anomaly = f\,(climate) + f\,(satellite) + county\ \ fixed\ \ effect \quad (1)$$

Here, "f" represents the fitting functions for climate and satellite predictors. The climate and satellite predictors are specified among monthly variables from the growing season (May to August), as shown in Table 1. For example, the predictor "VPD" represents a combination of four monthly VPD variables from May to August. The reason to choose model predictors at a monthly rather than a sub-monthly scale (e.g., weekly or daily) is to balance the model complexity, the risk of overfitting (Holzkämper et al., 2012), and the applicability with alternative climate data sources (Schauberger et al., 2017). The county fixed-effect accounts for location-specific and time-invariant (relatively) characteristics that are not reflected in climate/satellite data. These include different soil types, managements/practices, and seeds varieties, etc. Soil-property related variables (e.g., organic matter and soil available water content) were not explicitly included in the model because in our experiments they offer no help in improving the prediction. The county-fixed effect was estimated for each county separately by adding each unique "FIPS" code as a categorical variable in the model. Each county has its own intercept estimated in the model, which is equivalent of subtracting a group mean from observations of each county (Blanc and Schlenker, 2017). Note that additional interaction terms among climate or other variables can be incorporated as well. Interactions terms such as VPD × precipitation could potentially improve the prediction (Urban et al., 2015). However, our test showed that adding interaction terms VPD × P or T × P leads to only marginal (~1%) or no improvement for prediction. Therefore, in the following analysis, we concentrate on other modeling practices that significantly improve the prediction (i.e., predictor selection and fitting functions).

For each climate or satellite predictor, different fitting functions can be applied. Here we consider linear functions as well as nonlinear functions, including 2nd-order polynomials (i.e., quadratic) and 1st-order splines (i.e., piecewise linear spline). These different fitting functions lead to different fitted yield responses, which consequently affect model performance. Spline fitting differs from polynomial fitting as it requires more customization with specified "knots", whereas polynomial fitting is an automatic procedure that can be done without much human input.

To better understand the yield response, Fig. 1 shows scatterplots between yield anomaly and different climate and satellite predictors (columns) in different months (rows). The solid red line is produced by Locally Weighted Scatterplot Smoothing (LOWESS), which is a non-parametric smoothing method that can recover the underlying non-linear response from a raw scatter plot. LOWESS smoothing itself is not a prerequisite for building the model, but it helps identify knots which were used by the spline function to characterize the nonlinear relationship of a predictor. Apparently, yield response to climate-related predictors (VPD, air temperature, precipitation, LST) is nonlinear and varies by month, where crop yield declines beyond the optimal growth range. While EVI in the early growing season (May and June) has a flattened relationship with yield due to weak vegetation signal and contamination of bare soil, EVI during peak growing months (July and August in Fig. 1) shows a strong positive and almost linear relationship with yield.

To construct the model, the first step is to determine what predictor and fitting function to use. Different combinations of predictor and fitting function give multiple model configurations. For example, the model can be configured using either T + P, or VPD + P, or other combinations. Then it may be fit with a given function (linear, polynomial, or spline). These two models above with spline fitting can be written as $(VPD + P)_{spline}$ and $(T + P)_{spline}$, respectively. The "VPD" and "T" stand for four monthly VPD and temperature predictors during the growing season months from May to August as listed in Table 1, and "P" stands for precipitation in a similar way. The county-specific fixed-effect was included by default for all model configurations (see an exemplary map of the county fixed-effect in Fig. S1). Therefore, all model structures can be written as shown in Fig. 2.

Each model was trained using historical climate/satellite data including both spatial and temporal variability, and are therefore considered to be "panel" models (Lobell and Burke, 2010). The model coefficients were estimated by the least-squares method using the "stats" package in R. The predicted yield anomaly after adding the global yield trend gives the final predicted yield.

2.3. Model evaluation

Each model's yield prediction performance can be evaluated by an out-of-sample test, specifically, by either a "leave-one-year-out" (Mkhabela et al., 2011) or "forward" test. For the "leave-one-year-out" test, a model was first trained using the data with one year removed, and then the trained model was used to predict the yield of the missing year. This procedure was performed for each year from 2003 to 2016, resulting in a total of 14 years' yield predictions. The first prediction year was 2003 because both the MODIS satellite variables (LST and EVI) became available only after 2003. For each test year, the "leave-one-year-out" test uses the same amount of years for training, so the model performance is not affected by the length of training data. Therefore, the results reflect the model's overall prediction ability under various conditions. For this reason, "leave-one-year-out" is used as the default out-of-sample test for model evaluation throughout the paper unless otherwise stated.

For the "forward" test, only data before the test year were used to train the model and then make predictions for the test year. The "forward" test emulates the real-world yield prediction problem because only existing historical data can be used to train the model. However, model performance in this case would presumably increase over time as the prediction in later years has more training data than that of earlier years. The "forward" prediction was performed from 2005 to 2016 (a total of 12 years' predictions), and the results are presented for the regional averaged yield prediction in Section 3.5.

The model performance was evaluated with R^2 and Root-Mean-Square-Error (RMSE), which were calculated separately for each prediction year. R^2 here is defined as the square of the linear correlation coefficient between predicted and actual yield, while there are alternative definitions in the literature (see discussion in Section 4). Yield predictions with higher R^2 and lower RMSE are deemed to be better.

3. Results

3.1. Variable and fitting function selection

Since there are many possible model configurations, we present the modeling results in an order corresponding to each question raised at the end of the introduction. To determine the best combination of climate predictors (VPD, T, P) and their time scale (growing season, monthly), we compared models' performance among combinations of "growing season mean T and P" ($T_{gs} + P_{gs}$), "monthly T and P" (T + P), and "monthly VPD and P" (VPD + P), all with the same linear fitting function (Fig. 2a, b). For choice of climate variables, VPD appears to be a better predictor than T, because VPD in combination with precipitation (VPD + P) has a much higher predictive performance than the temperature model (T + P). The strong influence of VPD on maize growth has been attributed to the plant's stomatal response to VPD (Ball et al., 1987; Lobell et al., 2014, 2013). As for the time scale of predictors, the model built using climate variables (T + P) at a monthly scale has better performance (higher R^2 and lower RMSE) than the model at the growing season scale ($T_{gs} + P_{gs}$) (Fig. 2a, b). This is consistent with the findings of Holzkämper et al. (2012) that the disaggregated approach (i.e., monthly) is better than the aggregated approach (i.e., growing season) when using a large sample size.

To determine which fitting function leads to the best prediction, we compared models using climate predictors VPD + P, but configured with linear, polynomial, and spline fitting functions (Fig. 2c,d). Among different fitting functions, nonlinear functions (polynomial and spline) outperform linear functions because they can better capture the nonlinear yield response. Spline functions with specified knots show an overall higher R^2 and lower RMSE than polynomial functions. Therefore, VPD + P with a spline function, $(VPD + P)_{spline}$, is deemed as the "Best Climate-only" model in this study, with a median R^2 of 0.79, and RMSE of 1.04 t/ha from 2003 to 2016.

To understand the benefit of satellite variables to yield prediction, we compared various models constructed purely with satellite variables (Fig. 2e,f) and examined whether they can further improve prediction when combined with climate variables (Fig. 2g,h). Results show that yield prediction can be significantly improved with satellite variables. Models using either LST (LST_{spline}) or EVI alone (EVI_{linear}, EVI_{poly}) achieve R^2 and RMSE close to the "Best Climate-only" model. EVI in polynomial form leads to a slightly better prediction than EVI in the linear form, because the EVI–yield relationship is not strictly linear (Fig. 1). When including both LST and EVI as predictors, the resulting satellite model ($LST_{spline} + EVI_{poly}$) outperforms the "Best Climate-only" model by a large margin, with a median R^2 of 0.82 and RMSE of 0.97 t/ha from 2003 to 2016. Incorporating satellite variables into the "Best Climate-only" model leads to a higher performance than any models solely using climate or satellite variables (Fig. 2g,h). Among different configurations, $(VPD + P)_{spline} + EVI_{poly}$ stands out as the best of all models (hereafter referred to as the "Best Climate + EVI" model). It has a median R^2 of 0.85 and RMSE of 0.90 t/ha from 2003 to 2016, respectively (see each year's prediction in Fig. S2).

Fig. 3 shows the county-level temporal correlation of the predicted yield by the "Best Climate + EVI" model with actual yield (Fig. 3a), and

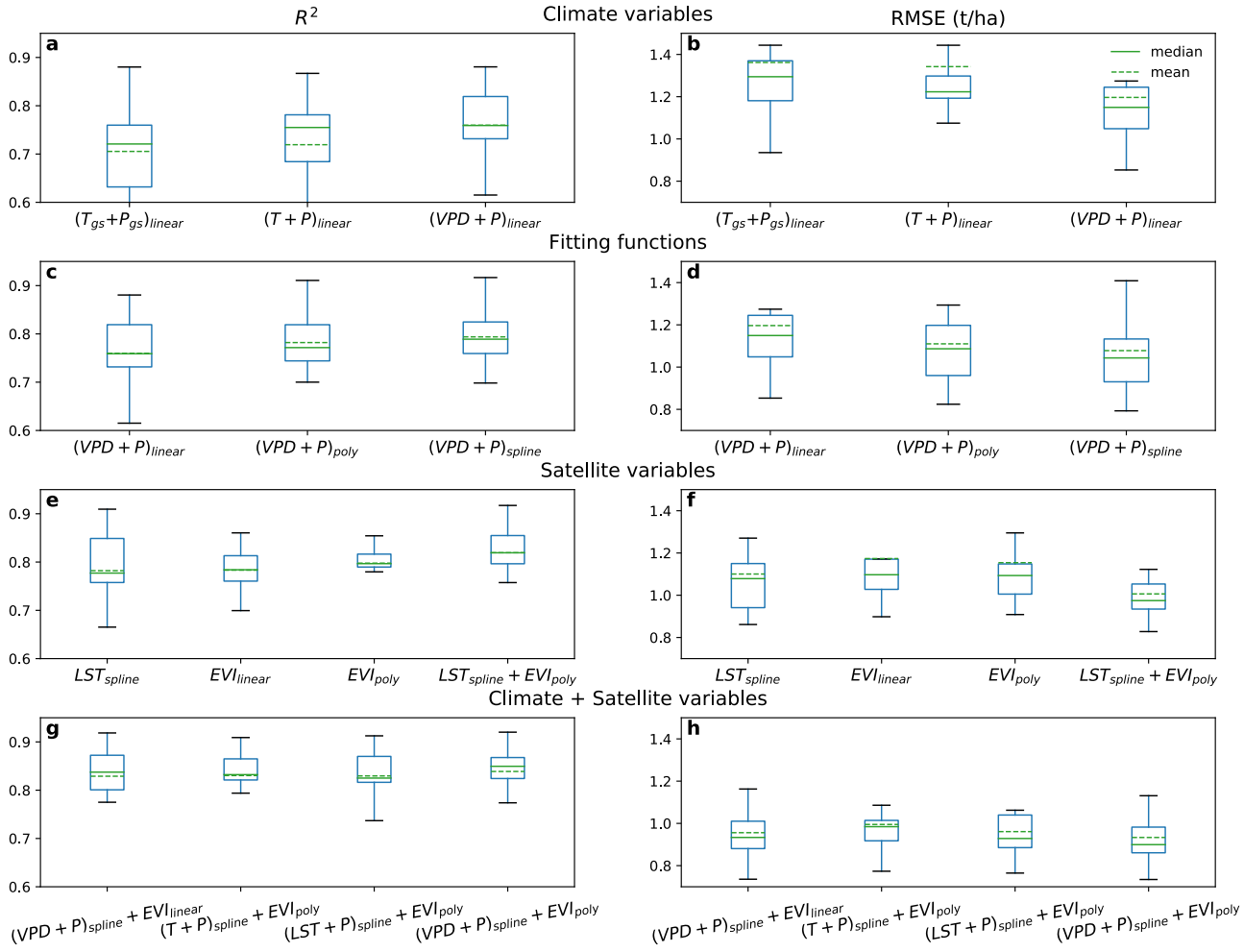

Fig. 2. Comparison of the predictive performance of models configured with different climate variables (a,b), different fitting functions (c,d), satellite variables alone (e,f), and with climate + satellite variables. Performance is evaluated by R^2 and RMSE from the "leave-one-year-out" test from 2003 to 2016. Climate variables include growing season mean air temperature (T_{gs}), precipitation (P_{gs}), monthly mean air temperature (T), VPD, and precipitation (P). Fitting functions include linear, polynomial (2nd-order), and spline (1st-order). Satellite variables include monthly EVI and max LST from MODIS. See full description of each model's configuration and its performance in Table S1.

their prediction RMSE (Fig. 3b). The yield prediction shows a high temporal correlation and low RMSE in most counties, particularly in the core part of the U.S. Corn Belt. However, large biases are found in some counties, suggesting that good performance is not guaranteed everywhere. For counties with large biases, we identified that their low correlation and high RMSE are associated with small harvested area and fewer years of yield data (Fig. S3). It is likely that the yield data in counties with small harvested area are collected from a relatively small sample size, and thus the yield is less representative and subject to larger uncertainty. Also, evaluating the model against fewer years of

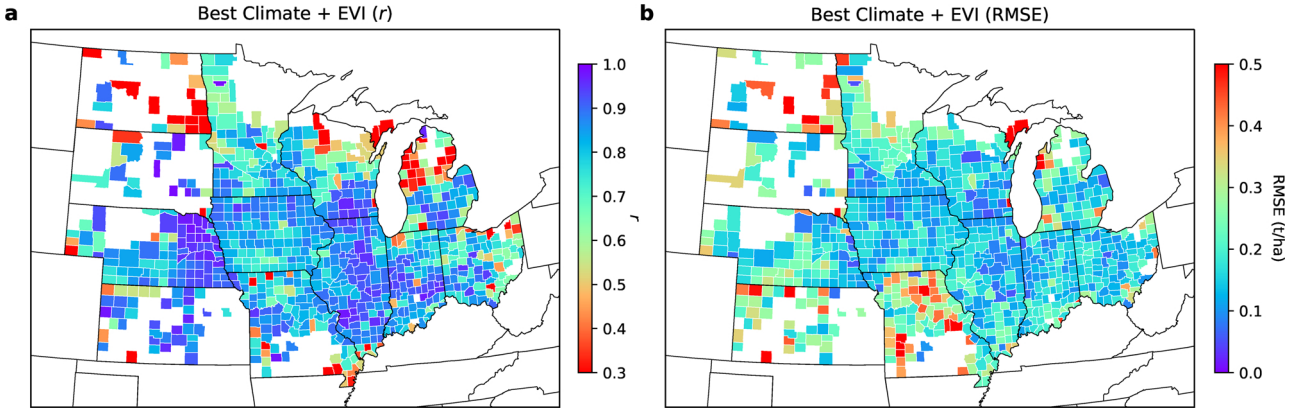

Fig. 3. (a) Temporal correlation between actual yield and predicted yield by the "Best Climate + EVI" model and (b) its median prediction RMSE from 2003 to 2016 at each county. Counties with yield observations fewer than 5 years over the evaluation period are not shown.

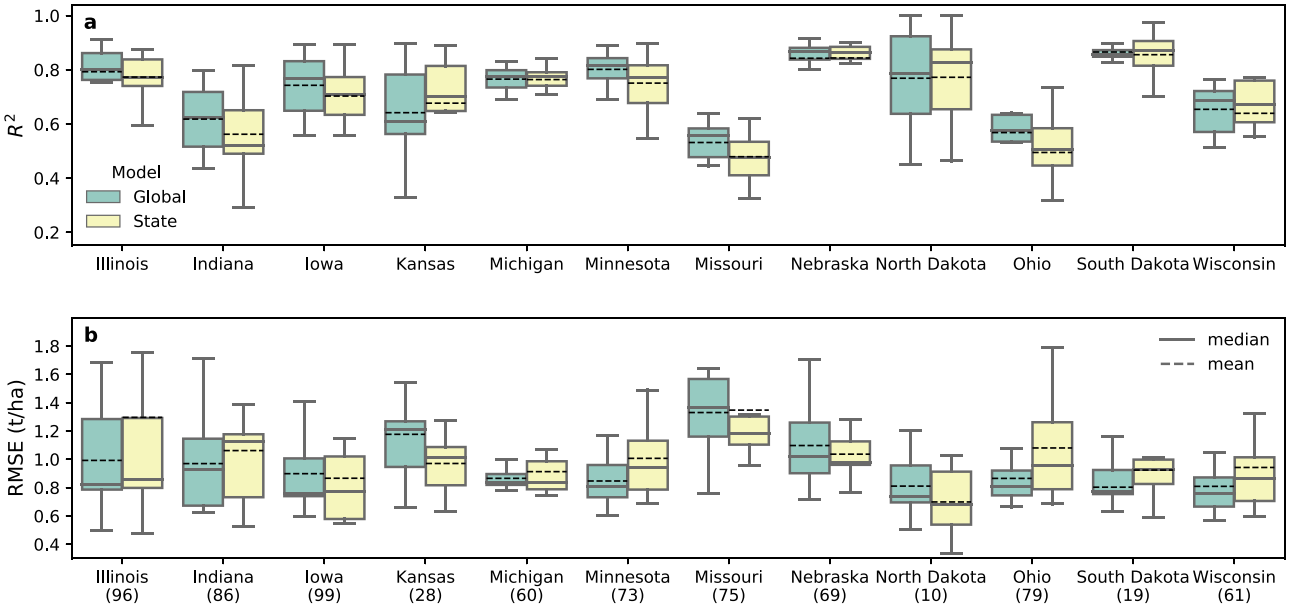

Fig. 4. Comparison of the predictive performance between the global model and state-specific models in different states. Panel (a) and (b) show the median R^2 and RMSE from 2007 to 2016 with the "leave-one-year-out" test for model $(VPD + P)_{poly} + EVI_{poly}$. The evaluation begins in 2007 because all states have EVI data after this year. The number in parenthesis is the mean county sample size of each state from 2007 to 2016.

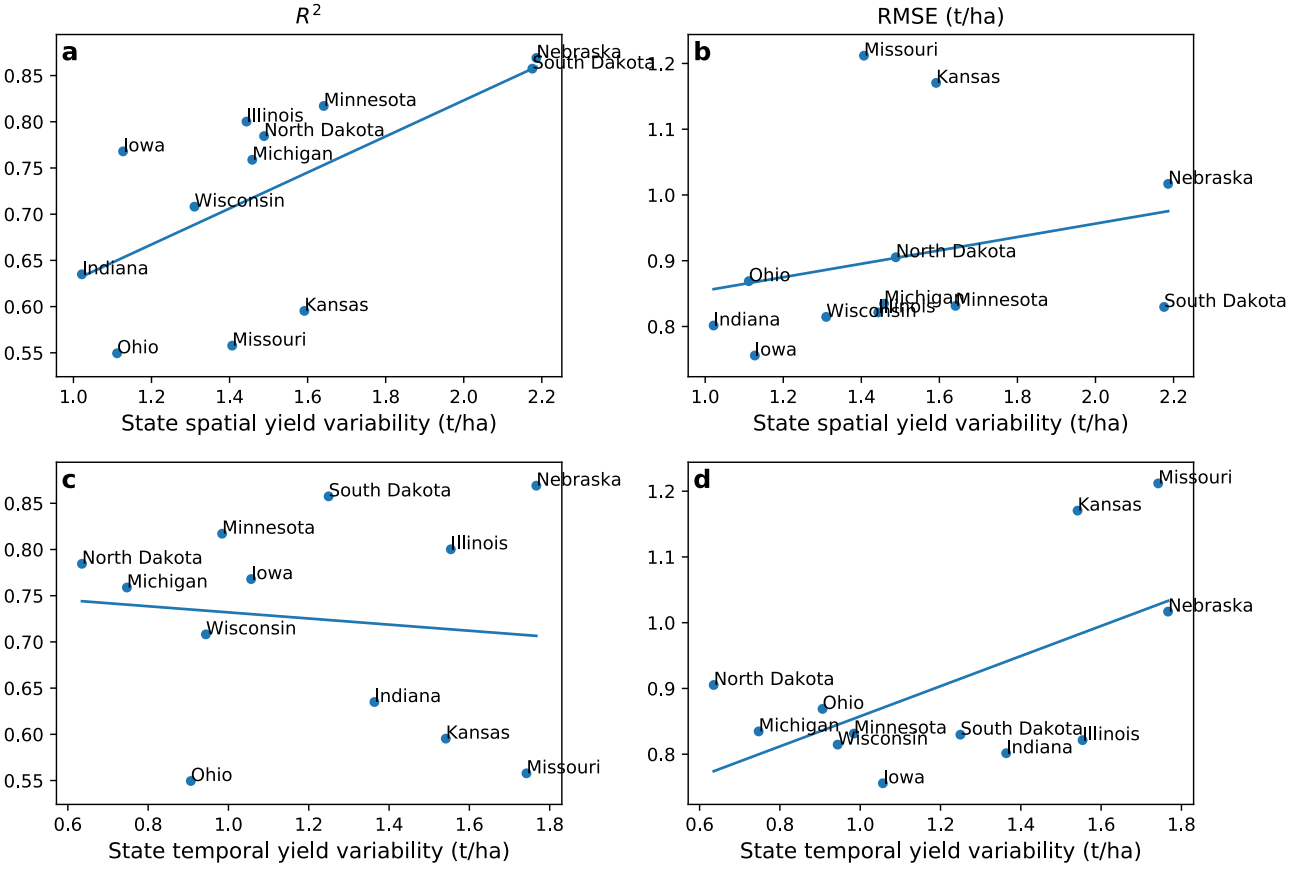

Fig. 5. The regional variations in models' predictive performance are related to (a,b) spatial and (c,d) temporal yield variability of a state. R^2 and RMSE are the median value for each state from 2003 to 2016 with model $(VPD + P)_{poly} + EVI_{poly}$. State spatial yield variability is calculated as the standard deviation of observed yield of each year averaged from 2003 to 2016. State temporal yield variability is calculated as the standard deviation of the state averaged yield time series from 2003 to 2016.

data is less robust and can be easily affected by interannual yield variability. Additionally, uncertainties in climate and satellite data, which vary spatially (Becker-Reshef et al., 2010), also contribute to degraded performance in certain regions.

3.2. Spatial scale to train the model

Statistical models of a given structure can be trained at different spatial scales, affecting their prediction performance. Two spatial scales are considered for model training: the global and state-specific models. The "global model" is a single model that uses data from all 12 states for training and assumes a universal yield relationship with climate and satellite variables. The state-specific model trains different models for each state by only using the data from that state. The rationale is that the "universal" yield–climate relationship could be altered by region-specific factors (e.g., climate regimes, soil conditions, management practice) and thus could differ in different states. To simplify implementation of the state-specific model, we used the model configuration $(VPD + P)_{poly} + EVI_{poly}$ because polynomial fitting can achieve a similar performance (only slightly lower, Table S1) to spline fitting without the need to specify knots for each state.

The performance of the global and state-specific models in different states is compared in Fig. 4. In most states, the global model has better (higher R^2 and lower RMSE) or equivalent performance than the state-specific model, with exceptions in states such as Kansas. On the one hand, better performance of the global model probably reflects the importance of training data to data-driven models, because the global model with a larger training dataset can cover a wider range of yield conditions than state-specific models. On the other hand, exceptions in Kansas where state-specific model gives a better prediction also seem to support a regionally varying climate-yield relationship which implies the comparison results could be region-dependent. Another important feature is the considerable regional variations in the models' performance seen in both global and state models. For example, Missouri has a low R^2 and high RMSE while the opposite is true for South Dakota. This suggests the predictability of yield could differ in different states. The causes of such regional variations will be investigated in the next section.

3.3. Regional variations in models' performance

Further analysis reveals that regional variations in the models' predictive performance, measured by R^2 and RMSE, are related to spatial and temporal yield variability of each state (Fig. 5). The positive relationship in Fig. 5a indicates that states with higher predictive R^2 are associated with larger spatial yield variability. A similar pattern has been reported elsewhere (Schauberger et al., 2017). Moreover, the number of county samples each state has may also contribute to this pattern, because the standard deviation of yield would be greater in states with fewer county samples ($r = -0.46$). The predictive RMSE, on the other hand, is influenced by temporal yield variability of each state, as shown by their positive relationship in Fig. 5d. This indicates that states with smaller temporal yield variability tend to show lower RMSE.

3.4. Interannual variations in models' performance

The models' performance exhibits considerable interannual variations, as demonstrated by predictions from 7 states by "Best Climate-only" and "Best Climate + EVI" model (Fig. 6). The 7 states subset ensures both models have yield predictions for as long as possible for interannual analysis (2003–2016), and the states included are the same for every year. Comparison between the two models confirms that satellite information added in the "Best Climate + EVI" model improves prediction considerably over the "Best Climate-only" model that relies on climate information alone. This is particularly true for the extreme drought year of 2012, whose RMSE was reduced from 1.81 to 1.06 t/ha after adding EVI into the model. However, we see no improvement in 2010, which is an extreme wet year that has the lowest R^2 and highest RMSE. Overall, the high resemblance between interannual variations of two models implies some common causes that affect the yield predictability of each year.

Our diagnostics found that wetness and spatial yield variability are closely linked to the model's interannual variations in performance. The degree of wetness is represented by the standard anomaly of growing season precipitation (May to August), averaged from all counties of each year. The spatial yield variability is represented by the standard deviation of county yield in each year. Results reveal a "U-shaped" curve between wetness and prediction performance, where model performance degrades toward both extreme dry and wet conditions (lower R^2 and larger RMSE; Fig. 7a,b), particularly toward wet years. The largest RMSE (1.22 t/ha) and lowest R^2 (0.61) are found in 2010, which is the wettest year in the evaluation period. In the extreme drought year 2012, prediction still shows high R^2 (~0.86) and moderate RMSE (1.06 t/ha), with the help of EVI. This suggests that yield in wetter years is difficult to predict by the "Best Climate + EVI Model", even with the added EVI information. Spatial yield variability of each year is another factor that affects model performance (Fig. 7c,d). An increase in spatial yield variability generally boosts R^2, which means years with a larger yield variability would have a higher prediction R^2 (Schauberger et al., 2017). However, prediction RMSE tends to increase as well, indicating a mismatch between evaluation metrics under large spatial yield variability. To test if these findings are sensitive to the sample size and alternative models, we repeated this analysis with the "Best Climate-only" model and extended the "leave-one-year-out" test back to 1981 which resulted in a larger test sample of 36 years (1981 to 2016). Similar results were obtained (Fig. S4). Thus, the robustness of our findings is confirmed.

3.5. Forward prediction of regional averaged yield

Model performance in previous sections is evaluated with the "leave-one-year-out" test. This is suitable to assess the robustness of the model structure under various conditions, but it is also fundamentally different from the prediction/forecast problem in the real world. In this section, we examine our model's ability to predict regional averaged yield with the "forward" test to emulate the real-world yield forecast problem. The regional averaged yield is obtained by aggregating the yield prediction of all counties. Two aggregation methods can be used to obtain the regional yield: unweighted arithmetic-mean (each county with an equal weight) and area-weighted mean (each county weighted by their harvested area). The regional averaged yield predictions from the "Best Climate-only" and "Best Climate + EVI" models are provided in Fig. 8. The regional averaged prediction of both models can well capture the interannual variability in actual yield, regardless of the aggregation method. Although the "Best Climate + EVI" model has an overall smaller RMSE (0.289 vs 0.307 t/ha) for unweighted regional yield, the prediction of the "Best Climate-only" model is practically closer to actual yield in most years, except in 2012 where yield loss is significantly underestimated (Fig. 8a). Interestingly, when it comes to area-weighted regional yield, the "Best Climate-only" model outperforms the "Best Climate + EVI" model (RMSE: 0.294 vs 0.324 t/ha; Fig. 8b). However, this does not mean the "Best Climate-only" model leads to better predictions, because its county-level predictions still have larger biases than the "Best Climate + EVI" model in the "forward" test (Fig. S5). The "better" regional yield prediction is due to the spatial aggregation process in which county prediction errors cancel out. These results have several important implications. First, the model that achieved high performance at a fine scale (e.g., county-level) may not necessarily become the best model at the aggregated scale (e.g., national level, see also in Bussay et al., 2015), because spatial aggregation changes the prediction error distribution. Second, the differences in model performance for regional yield predictions between

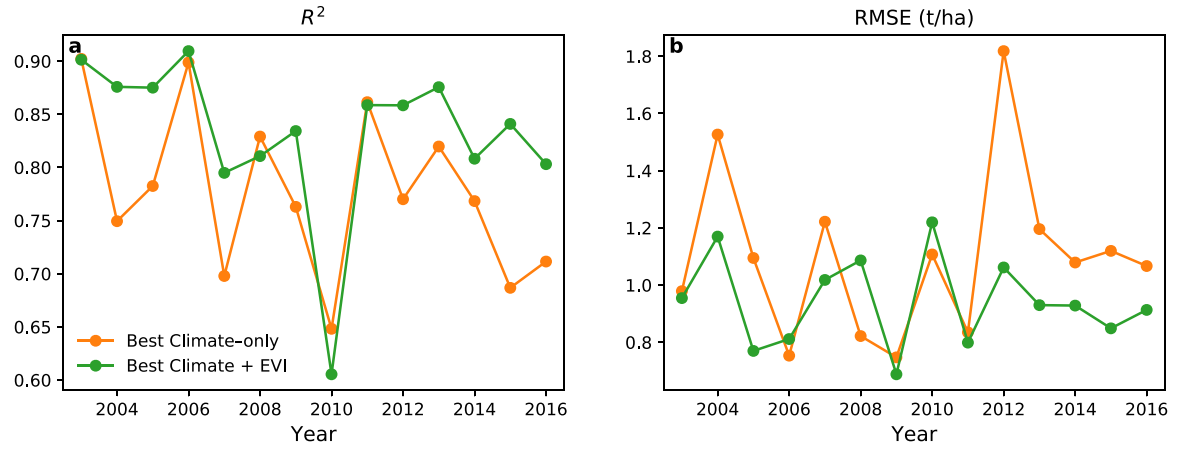

Fig. 6. Interannual variations in models' predictive performance for the "Best Climate-only" model and the "Best Climate + EVI" model. Note that only seven states (Illinois, Indiana, Iowa, Missouri, Nebraska, North Dakota, Wisconsin) with complete CDL data from 2003 and 2016 are included in this figure to avoid changes in state samples over time.

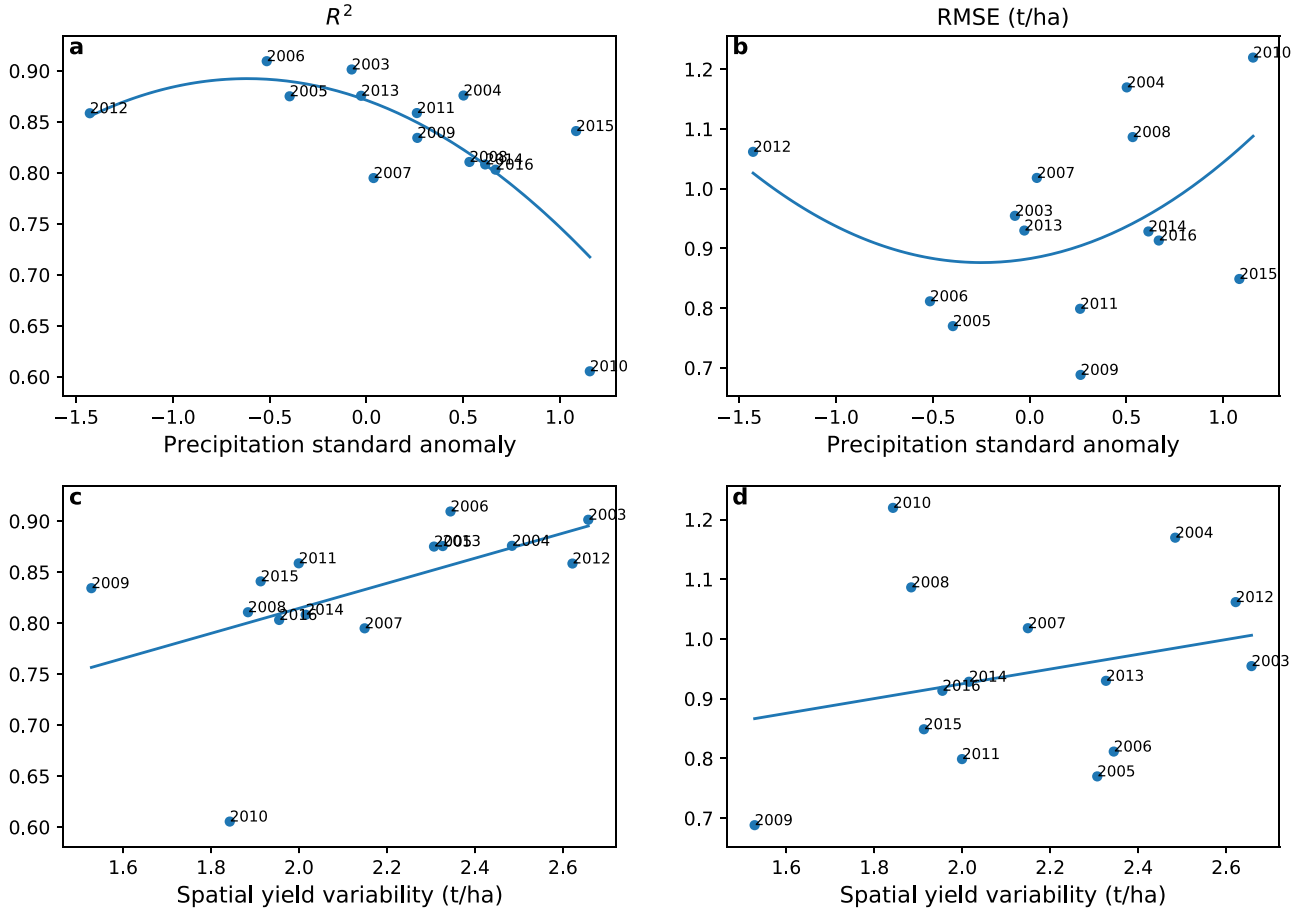

Fig. 7. Interannual variations in the predictive performance of the "Best Climate + EVI" model are related to wetness and spatial yield variability of each year. Wetness is represented by precipitation standard anomaly averaged of all counties. Positive and negative values represent wet and dry conditions, respectively. Spatial yield variability is calculated as the standard deviation of observed yield of each year. The fitted lines in panel (a) and (b) are 2-order polynomial and in panel (c), and (d) are linear. Only seven states (Illinois, Indiana, Iowa, Missouri, Nebraska, North Dakota, Wisconsin) with complete CDL data throughout 2003 and 2016 are included.

two aggregation methods highlight the effect of weighting. Weighing the yield prediction causes an unequal contribution of each county to the prediction error, which is not accounted for in the model training where each county is treated equally. We also tested assigning different weights to each county in model training but found no improvement. It is argued that yield weighting should not be used for model comparison as it adds another level of uncertainty (Schauberger et al., 2017).

Although the "Best Climate + EVI" model gives an overall better prediction than the "Best Climate-only" model, it needs more data input (satellite variables) and relies on CDL, which is usually released in the following year and thus needs additional efforts to estimate crop location (Cai et al., 2018) for real-time prediction. In this case, the "Best

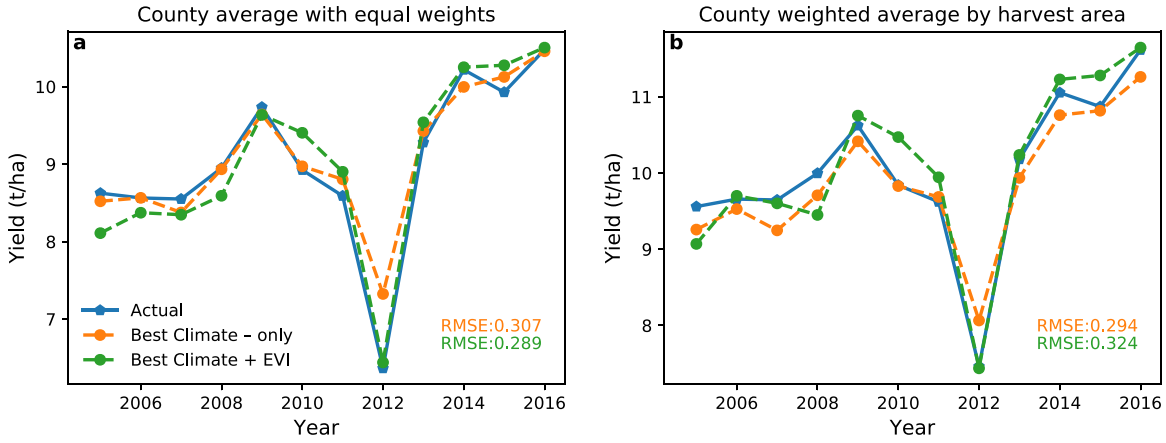

Fig. 8. Regional averaged yield prediction from the "forward" test by the "Best Climate-only" and "Best Climate + EVI" model from 2006 to 2016. (a) Unweighted regional yield prediction (arithmetic mean of county yield). (b) Area-weighted regional yield prediction (harvested area as weights). The comparison is constrained to counties that have a valid actual and predicted yield for both models. As a result, yield prediction prior to 2007 includes 7 states and afterward includes 12 states due to CDL data availability.

Climate-only" model still has its unique value for yield forecast. The model is purely based on climate information so it can directly utilize weather forecast information to predict yield with adequate leading time, even at the very early growing season.

4. Discussion

The modeling and diagnostic analysis presented in this study address several key issues in building a statistical model for crop yield prediction. Our results demonstrate that predictor and fitting function selection is critical for achieving higher predictive performance. Due to the data-driven nature of statistical models, model performance relies heavily on the training data. Our modeling practice for US rainfed corn shows that, in most cases, the model trained using the entire data (i.e., the global model) performs better than that using a subset of the data (Mathieu and Aires, 2016). However, exceptions exist in our results (Fig. 4) and in the literature (Gornott and Wechsung, 2016), suggesting its validity is likely to be region-dependent.

Our analyses highlight the great potential of incorporating satellite data in statistical crop yield modeling. We found that satellite information can greatly improve the model performance to a level that is not possible to reach by just using climate information alone (Balaghi et al., 2008; Becker-Reshef et al., 2010; Kern et al., 2018). This improvement is achieved by directly adding satellite-derived variables into the model, which is a relatively "easy" and "crude" approach. However, there are various other ways to make use of satellite data. For example, satellite information can be used to identify critical stages of crop growth (e.g., flowering, grain-filling for corn) during which the yield is particularly sensitive to certain environmental stress (Daryanto et al., 2016). This could potentially improve statistical crop models by developing the most effective weather predictor during certain phenological stages (as opposed to mean climate) for yield prediction (Bussay et al., 2015; Sakamoto et al., 2013). Also, there are many other crop- or environmentally-related variables that can be derived from different satellite sources or spectral bands, such as solar-induced fluorescence, backscattering, or vegetation optical depth, that can provide unique information to yield prediction (Guan et al., 2017b; Hank et al., 2018). But how to synthesize and make the best use of these satellite information for crop yield prediction is still an ongoing research question.

Through diagnostic analysis, we have demonstrated that the spatial and interannual variations in a model's predictive performance are closely linked to the spatial and temporal yield variability which in turn affects the predictability of yield. High spatial variability in yield boosts the predictive R^2 (Schauberger et al., 2017) but degrades the RMSE, whereas high temporal variability in yield degrades the RMSE, and vice versa. Although these diagnostics are mainly based on the "Best Climate + EVI" model, similar results were obtained with alternative model configurations such as $(LST + P)_{spline} + EVI_{poly}$, $(T + P)_{spline} + EVI_{poly}$, LST_{spline}, and EVI_{poly} (not shown), and with a longer evaluation period (Fig. S4). This consistency implies that these diagnostic results might be generalizable to the yield prediction of other statistical models. However, it is still unclear whether yield prediction by other independent approaches such as high-resolution satellite images, process-based models, or combined, would exhibit similar predictability features as statistical models.

In addition to variability in yield, extreme climate is another factor affecting yield predictability. The largest predictive RMSE of the "Best Climate + EVI" model and "Best climate-only" model appears in extreme dry (2012) and extreme wet years (2010), respectively. These unprecedentedly extreme climate events cause large prediction biases because such conditions may not have occurred in the historical training dataset. The prediction in extreme drought years can be greatly improved by incorporating satellite EVI, however, yield in wet years remains challenging to predict, which shows an increasing RMSE (decreasing R^2) as the year gets wetter (Fig. 7). This is probably because mechanisms of the drought impact on crop, such as high temperature and low moisture, can be directly and well captured by climate variables (e.g., temperature and precipitation). However, several key processes responsible for yield loss under wet conditions, such as root damage, delayed planting (Urban et al., 2015), lack of soil O_2 (Shaw and Meyer, 2015) are largely unaccounted by using climate data (rainfall) alone. Further, the impact of excessive rainfall on yield is more complicated than drought. Its yield impact can be either positive or negative, depending on the region and the timing of occurrence (Li et al., 2019), whereas drought impact is almost universally negative. All these additional complexities contribute to the challenge of yield prediction in wet years, and we call for more attention to the modeling of extreme climate impact.

The exemplary model presented in this study is for rainfed corn, but the modeling methodology itself is adaptable to irrigated corn and other crop types as well. Predicting irrigated yield differs from rainfed yield in at least two aspects. On the one hand, irrigation weakens the coupling between climate variability and yield, which in general leads to a decreased explanatory power of the model (decreased R^2). On the other hand, the low interannual variability of irrigated yield as a result of reduced sensitivity to climate variability also helps to reduce prediction RMSE. To demonstrate this, we applied our model to predict irrigated yield and combined yield (rainfed + irrigated). As expected, irrigated yield prediction has a smaller R^2 (0.44 vs 0.84) and lower RMSE (0.96 vs 1.07 t/ha) than rainfed yield, which translates to a

Table 2
Different model performance with different evaluation practices. Different definitions of R^2 and different ways of calculation lead to differences in performance. The model performance in the table is based on the "Best Climate-only" model with "leave-one-year-out" testing from 2003 to 2016.

Different R^2 definitions		Different methods to calculate R^2 and RMSE (2003-2016)				
R^2 (r × r)	R^2 (classic)	**Aggregate from individual years**		Lumping together all prediction years		
		R^2 (r × r)	**RMSE (t/ha)**	R^2 (r × r)	R^2 (classic)	RMSE (t/ha)
Median:0.789	Median 0.756	Median:0.789	Median: 1.043	0.821	0.809	1.103
Mean: 0.794	Mean: 0.762	Mean: 0.794	Mean:1.078			

Bold font shows the default choice of this study.

smaller R^2 (0.78) and a lower RMSE (0.91 t/ha) for combined yield (Fig. S6).

Although our statistical model achieved high yield prediction performance, it is worth mentioning the emerging trend of machine learning models in recent years, which has been experimentally used for yield prediction (e.g., Cai et al., 2017; Mathieu and Aires, 2016). Machine learning has the ability to discover useful information and uncover "hidden" features from a large set of historical or training data which could potentially lead to better prediction than traditional statistical methods. However, compared with statistical models, machine learning methods usually operate like a "black box" that lacks transparency and traceability. Therefore, statistical models will still play a key role in the yield prediction problem for their transparency, interpretability, simplicity, and relatively good performance.

Another issue worth discussing is the non-negligible influence of evaluation practices on model performance, specifically, the inconsistent definition of an evaluation metric (e.g., R^2) and different evaluation procedures (Table 2) (Richter et al., 2012). There are different definitions for the commonly used metric "R^2" in the literature. R^2, by its classic definition, is calculated as $1-SS_{res}/SS_{tot}$ (e.g., Mathieu and Aires, 2016), where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. But more recently R^2 has also been calculated as the square of linear correlation (r × r) between prediction and observation (e.g., Jin et al., 2017). For instance, the "Best Climate-only" model evaluated by R^2 has lower performance with the classic definition than with the "r × r" definition (0.76 vs. 0.79). Moreover, different evaluation procedures also lead to different model performances. Model performance can be evaluated by lumping all years' predictions together to calculate a single R^2/RMSE, or by aggregating the R^2/RMSEs calculated for each year's prediction individually. For the "Best Climate-only" model, the former approach gives a slightly higher R^2 (0.82 vs. 0.79) and larger RMSE (1.10 vs 1.00 t/ha) than the latter approach (the median of 2003–2016). The differences seem to be trivial, but they do matter for the incremental improvement of high-performance models. Finally, whether the prediction is made out-of-sample or in-sample, and which out-of-sample test to use (e.g., "leave-one-year-out" or "forward") also influences the evaluation results. All these nuances indicate that model evaluation is not as simple as it seems (Richter et al., 2012). The reported R^2 and RMSE, which are indicators of model performance, can be quite sensitive to some unintended choice in the evaluation process. This makes it very difficult to directly compare model performance across different studies, and this is further complicated by the spatial and temporal variations in the models' performance. The comparability issue could be partially alleviated by adopting a clearly-defined evaluation metric and a transparent evaluation procedure, but a more ideal solution is to make the prediction data open access so that it can be reanalyzed by a consistent methodology.

5. Conclusion

The modeling practices in this study pave the way for future development and application of statistical yield models. The exemplary model proposed here is for rainfed corn in the Midwest U.S. but it is generalizable for irrigated corn and other crop types (e.g. soybean, wheat, cotton, rice) in other regions. The methodology is flexible enough to ingest new data from either climate or satellite sources, as well as including additional interaction terms. Such models can be applied to within-season crop yield forecasting (Peng et al., 2018b) or climate change impact assessment in the near future. To further improve yield prediction, we emphasize the importance of performing diagnostic analyses of existing models, as well as applying a transparent model evaluation procedure and open data practice. The model diagnosis is not only needed to fully understand the strength of existing models, but also to identify their limiting factors, the underperformed conditions, and the possible causes. A transparent model evaluation procedure, and open data practices could benefit model development through improving the comparability of relevant studies and also make it possible to establish a model benchmark for cross comparison.

Acknowledgements

Y. L., K. G., A.Y., and B.P. acknowledge the supports from USDA NIFA (2017-68002-26789, 2017-67003-28703, and ILLU-875-953). Y.L. and K.G. designed the study, Y.L., K.G., and A.Y., performed the analysis. Y.L. and K.G. wrote the first draft of the manuscript with edits from A.Y., B.P., L.Z., B.L., and J.P. All authors contributed to the development of the statistical model. We thank two reviewers and the Editor for their insightful comments that significantly improved the manuscript. The yield prediction data of this study are available at Mendeley Data (doi:10.17632/hyf67hrnx2.1).

References

Anderson, M.C., Hain, C., Wardlow, B., Pimstein, A., Mecikalski, J.R., Kustas, W.P., 2011. Evaluation of drought indices based on Thermal remote sensing of evapotranspiration over the continental United States. J. Clim. 24, 2025–2044. https://doi.org/10.1175/2010JCLI3812.1.

Balaghi, R., Tychon, B., Eerens, H., Jlibene, M., 2008. Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. Int. J. Appl. Earth Obs. Geoinf. 10, 438–452. https://doi.org/10.1016/j.jag.2006.12.001.

Ball, J.T., Woodrow, I.E., Berry, J.A., 1987. A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. Progress in Photosynthesis Research. https://doi.org/10.1007/978-94-017-0519-6_48.

Bassu, S., Brisson, N., Durand, J.L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S.H., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Glob. Change Biol. 20, 2301–2320. https://doi.org/10.1111/gcb.12520.

Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens. Environ. 114, 1312–1323. https://doi.org/10.1016/j.rse.2010.01.010.

Blanc, E., Schlenker, W., 2017. The use of panel models in assessments of climate impacts on agriculture. Rev. Environ. Econ. Policy 11, 258–279. https://doi.org/10.1093/reep/rex016.

Bussay, A., van der Velde, M., Fumagalli, D., Seguini, L., 2015. Improving operational maize yield forecasting in Hungary. Agric. Syst. 141, 94–106. https://doi.org/10.1016/j.agsy.2015.10.001.

Butler, E.E., Huybers, P., 2012. Adaptation of US maize to temperature variations. Nat.

Clim. Change 3, 68–72. https://doi.org/10.1038/nclimate1585.
Cai, Y., Moore, K., Elhaddad, A., Lessel, J., Townsend, C., Solak, H., Semret, N., Pellegrini, A., 2017. Crop Yield Predictions - High Resolution Statistical Model for Intra-season Forecasts Applied to Corn in the US. Accessed on 24 December, 2018 at. https://gro-intelligence.com/yield-model/eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJhY2Nlc3MiOiJ5aWVsZE1vZGVsaW5nIiwiaWF0IjoxNTEwMjY3ODE3fQ.kpkw6shkDqD-Xzv58hyQ8DDxiH2wfG4NSHyNPypNOZg.
Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.02.045.
Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. https://doi.org/10.1002/joc.1688.
Daryanto, S., Wang, L., Jacinthe, P.A., 2016. Global synthesis of drought effects on maize and wheat production. PLoS One 11, 1–15. https://doi.org/10.1371/journal.pone.0156362.
Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sens. Environ. 188, 9–25. https://doi.org/10.1016/j.rse.2016.11.004.
Gornott, C., Wechsung, F., 2016. Statistical regression models for assessing climate impacts on crop yields: a validation study for winter wheat and silage maize in Germany. Agric. For. Meteorol. 217, 89–100. https://doi.org/10.1016/j.agrformet.2015.10.005.
Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B., 2016. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. Glob. Change Biol. 22, 716–726. https://doi.org/10.1111/gcb.13136.
Guan, K., Sultan, B., Biasutti, M., Baron, C., Lobell, D.B., 2017a. Assessing climate adaptation options and uncertainties for cereal systems in West Africa. Agric. For. Meteorol. 232, 291–305. https://doi.org/10.1016/j.agrformet.2016.07.021.
Guan, K., Wu, J., Kimball, J.S., Anderson, M.C., Frolking, S., Li, B., Hain, C.R., Lobell, D.B., 2017b. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote Sens. Environ. 199, 333–349. https://doi.org/10.1016/j.rse.2017.06.043.
Guan, K., Li, Z., Gao, N., Feng, G., 2018. Using MODIS-Landsat fused data and ALOS-2/PARSAR-2 to map paddy rice and estimate crop yield for Thai Binh Province in Viet Nam. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 2238–2252. https://doi.org/10.1109/JSTARS.2018.2834383.
Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, Ja, Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.-E., Moran, M.S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., Griffis, T.J., 2014. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. Proc. Natl. Acad. Sci. U. S. A. 111, E1327–33. https://doi.org/10.1073/pnas.1320008111.
Hank, T.B., Berger, K., Bach, H., Clevers, J.G.P.W., Gitelson, A., Zarco-Tejada, P., Mauser, W., 2018. Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges, Surveys in Geophysics. Springer, Netherlands. https://doi.org/10.1007/s10712-018-9492-0.
Heft-Neal, S., Lobell, D.B., Burke, M., 2017. Using remotely sensed temperature to estimate climate response functions. Environ. Res. Lett. 12, 014013. https://doi.org/10.1088/1748-9326/aa5463.
Holzkämper, A., Calanca, P., Fuhrer, J., 2012. Statistical crop models: predicting the effects of temperature and precipitation changes. Clim. Res. 51, 11–21. https://doi.org/10.3354/cr01057.
Jin, Z., Azzari, G., Lobell, D.B., 2017. Improving the accuracy of satellite-based high-resolution yield estimation: a test of multiple scalable approaches. Agric. For. Meteorol. 247, 207–220. https://doi.org/10.1016/j.agrformet.2017.08.001.
Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens. Environ. 141, 116–128. https://doi.org/10.1016/j.rse.2013.10.027.
Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Brief history of agricultural systems modeling. Agric. Syst. 155, 240–254. https://doi.org/10.1016/j.agsy.2016.05.014.
Karnieli, A., Agam, N., Pinker, R.T., Anderson, M., Imhoff, M.L., Gutman, G.G., Panov, N., Goldberg, A., 2010. Use of NDVI and land surface temperature for drought assessment: merits and limitations. J. Clim. 23, 618–633. https://doi.org/10.1175/2009JCLI2900.1.
Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. Agric. For. Meteorol. 260–261, 300–320. https://doi.org/10.1016/j.agrformet.2018.06.009.
Li, Y., Guan, K., Schnitkey, G., Evan, D., Peng, B., 2019. Excessive rainfall leads to comparable magnitude of corn yield loss as extreme drought in the US. Glob. Chang. Biol (In revision).
Lobell, D.B., Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. Agric. For. Meteorol. 150, 1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008.
Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing climate change adaptation needs for food security in 2030. Science 319, 607–610. https://doi.org/10.1126/science.1152339.
Lobell, D.B., Hammer, G.L., McLean, G., Messina, C., Roberts, M.J., Schlenker, W., 2013. The critical role of extreme heat for maize production in the United States. Nat. Clim. Change 3, 497–501. https://doi.org/10.1038/nclimate1832.
Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M., Hammer, G.L., 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. Science 344, 516–519. https://doi.org/10.1126/science.1251423.
Mathieu, J., Aires, F., 2016. Statistical weather-impact models: an application of neural networks and mixed effects for corn production over the United States. J. Appl. Meteorol. Climatol. 55, 2509–2527. https://doi.org/10.1175/JAMC-D-16-0055.1.
Mkhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y., 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric. For. Meteorol. 151, 385–393. https://doi.org/10.1016/j.agrformet.2010.11.012.
Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R.C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T.A.M., Ray, D.K., Reddy, A., Rosenzweig, C., Ruane, A.C., Sakurai, G., Schmid, E., Skalsky, R., Song, C.X., Wang, X., De Wit, A., Yang, H., 2017. Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. Geosci. Model Dev. 10, 1403–1422. https://doi.org/10.5194/gmd-10-1403-2017.
Peng, B., Guan, K., Chen, M., Lawrence, D.M., Pokhrel, Y., Suyker, A., Arkebauer, T., Lu, Y., 2018a. Improving maize growth processes in the community land model: implementation and evaluation. Agric. For. Meteorol. 250–251, 64–89. https://doi.org/10.1016/j.agrformet.2017.11.012.
Peng, B., Guan, K., Pan, M., Li, Y., 2018b. Benefits of seasonal climate prediction and satellite data for forecasting U.S. Maize yield. Geophys. Res. Lett. 45, 9662–9671. https://doi.org/10.1029/2018GL079291.
Richter, K., Atzberger, C., Hank, T.B., Mauser, W., 2012. Derivation of biophysical variables from Earth observation data: validation and statistical measures. J. Appl. Remote Sens. 6https://doi.org/10.1117/1.JRS.6.063557. 063557-1.
Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T., a, M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. U. S. A. 111, 3268–3273. https://doi.org/10.1073/pnas.1222463110.
Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. Remote Sens. Environ. 131, 215–231. https://doi.org/10.1016/j.rse.2012.12.017.
Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2014. Near real-time prediction of U. S. Corn yields based on time-series MODIS data. Remote Sens. Environ. 147, 219–231. https://doi.org/10.1016/j.rse.2014.03.008.
Schauberger, B., Gornott, C., Wechsung, F., 2017. Global evaluation of a semiempirical model for yield anomalies and application to within-season yield forecasting. Glob. Change Biol. 23, 4750–4764. https://doi.org/10.1111/gcb.13738.
Schlenker, W., Lobell, D.B., 2010. Robust negative impacts of climate change on African agriculture. Environ. Res. Lett. 5, 014010. https://doi.org/10.1088/1748-9326/5/1/014010.
Schlenker, W., Roberts, M.J., 2006. Nonlinear effects of weather on corn yields. Rev. Agric. Econ. 28, 391–398. https://doi.org/10.1111/j.1467-9353.2006.00304.x.
Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to U.S. Crop yields under climate change. Proc. Natl. Acad. Sci. 106, 15594–15598. https://doi.org/10.1073/pnas.0906865106.
Shaw, R.E., Meyer, W.S., 2015. Improved empirical representation of plant responses to waterlogging for simulating crop yield. Agron. J. 107, 1711–1723. https://doi.org/10.2134/agronj14.0625.
Shi, W., Tao, F., Zhang, Z., 2013. A review on statistical models for identifying climate contributions to crop yields. J. Geogr. Sci. 23, 567–576. https://doi.org/10.1007/s11442-013-1029-3.
Siebert, S., Ewert, F., Rezaei, E.E., Kage, H., Graß, R., 2014. Impact of heat stress on crop yield—on the importance of considering canopy temperature. Environ. Res. Lett. 9, 044012. https://doi.org/10.1088/1748-9326/9/4/044012.
Tack, J., Barkley, A., Nalley, L.L., 2015. Effect of warming temperatures on US wheat yields. Proc. Natl. Acad. Sci. 112, 201415181. https://doi.org/10.1073/pnas.1415181112.
Urban, D., Roberts, M.J., Schlenker, W., Lobell, D.B., 2012. Projected temperature changes indicate significant increase in interannual variability of U.S. Maize yields: a Letter. Clim. Change 112, 525–533. https://doi.org/10.1007/s10584-012-0428-2.
Urban, D.W., Roberts, M.J., Schlenker, W., Lobell, D.B., 2015. The effects of extremely wet planting conditions on maize and soybean yields. Clim. Change 130, 247–260. https://doi.org/10.1007/s10584-015-1362-x.
Wang, K., 2014. Sampling biases in datasets of historical mean air temperature over land. Sci. Rep. 4, 4637. https://doi.org/10.1038/srep04637.
Zeng, L., Wardlow, B.D., Wang, R., Shan, J., Tadesse, T., Hayes, M.J., Li, D., 2016. Remote Sensing of Environment A hybrid approach for detecting corn and soybean phenology with time-series MODIS data. Remote Sens. Environ. 181, 237–250. https://doi.org/10.1016/j.rse.2016.03.039.